



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

The heroism paradox

Citation for published version:

Archer, A & Ridge, M 2015, 'The heroism paradox: another paradox of supererogation', *Philosophical Studies*, vol. 172, no. 6, pp. 1575-1592. <https://doi.org/10.1007/s11098-014-0365-1>

Digital Object Identifier (DOI):

[10.1007/s11098-014-0365-1](https://doi.org/10.1007/s11098-014-0365-1)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Philosophical Studies

Publisher Rights Statement:

© Archer, A., & Ridge, M. (2015). The heroism paradox: another paradox of supererogation. *Philosophical Studies*, 172(6), 1575-1592. 10.1007/s11098-014-0365-1 / The final publication is available at Springer via <http://dx.doi.org/10.1007/s11098-014-0365-1>

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



The Heroism Paradox

Another Paradox of Supererogation

Philosophers are by now familiar with “the” paradox of supererogation. This paradox arises out of the idea that it can never be permissible to do something morally inferior to another available option, yet acts of supererogation seem to presuppose this.¹ This paradox is *not* our topic in this paper. We mention it only to set it to one side and explain our subtitle. In this paper we introduce and explore *another* paradox of supererogation, one which also deserves serious philosophical attention.

People who perform paradigmatic acts of supererogation very often claim and believe that their acts were obligatory. Plausibly, this is simply a mistake insofar as the actions really are “above and beyond the call of duty,” as common sense would have it. The fact that moral heroes tend to view their actions in this apparently mistaken way is puzzling in itself, and we might learn something interesting about the moral psychology of such individuals if we could explain this tendency. However, this puzzling aspect of the moral psychology of moral heroes is also the chief ingredient in a deeper puzzle, one perhaps more worthy of the title “paradox.” In this paper we present and try to resolve this paradox.

The paradox arises when we combine our initial observation about the moral psychology of moral heroes with three plausible claims about how these cases compare with one in which the agent realizes her act is “above and beyond.” The first of these three additional claims is that the agent who mistakenly claims that the act is obligatory

¹ For attempts to solve this paradox see Ferry (2013), Horgan and Timmons (2010) and Portmore (2008).

is no less virtuous than someone who performs such an act whilst correctly judging it to be obligatory. The second is that the agent who makes such a mistake would display more moral wisdom if she judged the act to be supererogatory. The third is that there is no other relevant difference between the two agents. These three claims, together with a plausible principle about the way in which the virtues work, give rise to a paradox. We consider several ways in which this paradox might be resolved. We argue that the most plausible resolution is to reject the claim that there is no other relevant difference between the two agents. More specifically, we argue that a relevant difference is that the agent who makes this mistake possesses a greater degree of moral depth.

1. The Phenomenon

Supererogatory acts are those that are beyond the call of duty. These include acts of tremendous heroism or saintliness that go beyond what anyone could reasonably think of as being morally required. An interesting phenomenon about people who perform such acts is that they often claim to have been merely doing their duty.²

Take the actions for which Daniel Inouye received a Medal of Honour. These acts seem like an obvious example of supererogation. The citation for Inouye's Medal of Honour reads as follows:

Second Lieutenant Daniel K. Inouye distinguished himself by extraordinary heroism in action on 21 April 1945, in the vicinity of San Terenzo, Italy. While attacking a defended ridge guarding an important road junction, Second Lieutenant Inouye skillfully directed his platoon through a hail of automatic weapon and small arms fire, in a swift enveloping movement that resulted in the capture of an artillery and mortar post and brought his men to within 40 yards

² This phenomenon is identified by Carbonell (2012), Driver (2001) and Heyd (2011).

of the hostile force. Emplaced in bunkers and rock formations, the enemy halted the advance with crossfire from three machine guns. With complete disregard for his personal safety, Second Lieutenant Inouye crawled up the treacherous slope to within five yards of the nearest machine gun and hurled two grenades, destroying the emplacement. Before the enemy could retaliate, he stood up and neutralized a second machine gun nest. Although wounded by a sniper's bullet, he continued to engage other hostile positions at close range until an exploding grenade shattered his right arm. Despite the intense pain, he refused evacuation and continued to direct his platoon until enemy resistance was broken and his men were again deployed in defensive positions. In the attack, 25 enemy soldiers were killed and eight others captured. By his gallant, aggressive tactics and by his indomitable leadership, Second Lieutenant Inouye enabled his platoon to advance through formidable resistance, and was instrumental in the capture of the ridge. Second Lieutenant Inouye's extraordinary heroism and devotion to duty are in keeping with the highest traditions of military service and reflect great credit on him, his unit, and the United States Army.³

Despite his horrendous injuries, Inouye chose to remain on the battlefield to assist his platoon. It seems hard to imagine anyone denying that Inouye went beyond the call of duty in acting as he did. However, one person who did deny this was Inouye himself. In an interview in 2010 Inouye said the following: "The pain was nothing great, and I had my job, my obligation and my mission to accomplish."⁴ If we take Inouye at his word then clearly he views his act as being obligatory not supererogatory.

³ Gomez-Granger (2008 p.11).

⁴ Boghani (2012)

By itself Inouye's assessment of his actions is interesting. It is strange that Inouye's assessment of his own acts differs from that of most other people. Even more interesting, though, is that the conflict between Inouye's assessment and that of others is not unique. In their study of moral exemplars Anne Colby and William Damon observed that people who dedicate their lives to moral causes often feel that they are obliged to do so (1992 p.70). For example, Suzie Valdez, who has dedicated her life to feeding the poor of Ciudad Juarez, made it clear to Colby and Damon that she felt she had to help in this way (1992 p.45). Similarly the civil rights activist Virginia Durr claimed that when it came to considering whether or not to dedicate her life to this cause "there were no choices to make" (1992 p.70).⁵ Inouye's case, then, is just one example of a more general phenomenon. Often people who perform supererogatory actions insist that they were only doing their duty.

We might think that witnessing this phenomenon should give us reason to question our judgment that the act in question is supererogatory. Vanessa Carbonell pursues this line of thought in a recent paper (2012). Carbonell argues that this phenomenon leads to a 'ratcheting-up effect' on the level of obligation faced by everyone else. Saints and heroes have a special insight into exactly what it is like to make the kind of sacrifices that they do. Carbonell claims that the fact that it is the people with this special insight that take their acts to be morally obligatory is important. It gives us reason to think that what we can reasonably demand of each other might be higher than previously thought (2012 p.238). Nothing we say in this paper contradicts Carbonell's claim that the behavior of

⁵ Other examples of people who deny the seemingly obvious supererogatory status of their acts include 'Subway Hero' Wesley Autrey who risked his life to save the life of a man who had fallen onto the track of the Subway. After the incident Autrey told the New York Times, "I don't feel like I did something spectacular; I just saw someone who needed help. I did what I felt was right," (Buckley 2007). Similarly, William Harvey Carney, awarded a Medal of Honour for acts of bravery in battle, declared afterwards, "Boys, I only did my duty," (Mulderink 2012 p.110).

moral saints might have this effect on how we should view our moral obligations. For all we say here it may well be the case that sometimes the correct response to witnessing this phenomenon is to revise our view of what is morally obligatory. Nevertheless, we take it that there are cases, such as Daniel Inouye's, in which the agent is mistaken about the deontic status of the action.

We might also think that the agent's ignorance in cases such as these is a kind of moral virtue. In fact, Julia Driver has claimed that this phenomenon provides support for her claim that some moral virtues are partly constituted by ignorance about certain relevant facts (2001: 40-41).⁶ Driver takes this to show that an acceptable account of the morally virtuous agent must tolerate that agent being ignorant or mistaken in certain ways. This is an interesting claim and what we have to say will provide a partial vindication of Driver's thesis. However, we will only be in a position to show why this is the case once we have introduced our paradox and our solution to it. For now, then, we will set this issue to one side.

2. Three Claims

In this section we put forward three intuitively plausible claims about the character of people who, like Daniel Inouye, deny the supererogatory status of their acts. These claims are comparative. To make the relevant comparison we must consider an alternate version of Inouye. In particular, we are interested in the counterfactual question, "would Inouye be more, less or equally virtuous if he did not make this mistake." Consider the nearest possible world (we initially assume there is one unique such world; we revisit this simplifying assumption below) in which Inouye both performs the action and judges it to be supererogatory, where 'supererogatory' is now

⁶ Thanks to an anonymous referee for bringing Driver's discussion of the phenomenon to our attention.

just a technical term intended to refer to whatever feature or cluster of features is picked out by ordinary language idioms like “above and beyond the call of duty.” Call the alternate version of Inouye that inhabits this world “Wise Inouye.” With this comparison case explained, we are in a position to put forward the three claims.

Claim 1 [“Not More Virtuous”]: Wise Inouye is not more morally virtuous overall than Inouye (and might even be less virtuous overall).

The first claim to make about the comparison between Inouye and Wise Inouye is that Inouye does not seem to be more virtuous overall than Wise Inouye. The fact that Wise Inouye is right about the moral status of the act would not lead us to judge the overall character of Wise Inouye more positively. In fact we might think that Inouye is more virtuous than Wise Inouye. As we have already seen, Inouye insisted in an interview that his act was obligatory. Imagine witnessing the equivalent interview with Wise Inouye who, when asked, says instead that he acted in a way that was above and beyond the call of duty. When we compare these two cases we think that many would share the thought that Inouye is more virtuous than Wise Inouye. This, though, is the stronger version of this claim and is not necessary for the argument we will make in the remainder for this paper. All we need to establish here is the weaker claim that Wise Inouye is not more virtuous overall than Inouye.

Claim 2 [“More Moral Wisdom”]: Wise Inouye possesses more moral wisdom (itself a moral virtue) than Inouye.

The second claim to make is that Wise Inouye has a higher level of moral wisdom than Inouye. This claim seems to follow straightforwardly from the fact that, unlike Inouye, Wise Inouye has correctly judged that his act was supererogatory rather than obligatory. Wise Inouye has shown that he is a better judge of the moral status of actions than

Inouye and so it is right to view Wise Inouye as possessing more moral wisdom. Of course, this will only be the case if the other differences that exist between Inouye and Wise Inouye are not ones that impact negatively on the amount of moral wisdom possessed by Wise Inouye. However, there seems no good reason to think that the closest possible world that meets our criteria will be one in which Wise Inouye has less moral wisdom than Inouye. In the absence of such a reason, we are safe to assume that Wise Inouye does possess more moral wisdom than Inouye.

More controversial here is the claim that moral wisdom is a moral virtue. It might be objected that moral wisdom is an epistemic virtue not a moral virtue.⁷ This claim has some prima facie appeal. After all, we might think of moral wisdom as a disposition to possess true and only (or mainly) true moral beliefs. Given this, it seems reasonable to class moral wisdom as an epistemic virtue.

In order to respond to this objection we need to have some way of distinguishing moral and epistemic virtues. This is a controversial area and we do not wish to commit ourselves to any specific account of the distinction. It is generally accepted, though, that what distinguishes epistemic virtues from moral virtues is that the former promote epistemic ends, such as truth and justified belief, while the latter promote moral ends.⁸ This distinction is quiet on exactly what should be considered a moral end and again we wish to say nothing controversial here. We take it that it is fairly uncontroversial to think that beneficence, autonomy and justice are all moral ends. We also take it that there is a fairly clear intuitive divide between these ends and epistemic ends.

⁷ Thanks to X and Y for raising this objection.

⁸ See Driver (2003 p.374), Montmarque (1986 p.488) and Zagzebski (1996 p.166-7). Of course, what we have said here is quiet on the way in which these virtues promote these different ends and the three accounts mentioned here provide different answers to that question. For our purposes though, we take it to be sufficient to point out that moral virtues promote moral ends and epistemic virtues promote epistemic ends.

Having said something about the distinction between moral and epistemic virtues it should be clear that moral wisdom is an epistemic virtue. It is a disposition to possess true and only true moral beliefs and so is clearly connected to epistemic ends.

However, this is only problematic for our claim if we think that moral wisdom cannot *also* be a moral virtue. The point we wish to make is that moral wisdom is a moral virtue. If it is also an epistemic virtue then this provides no problems for this claim. All we need to establish is that it is a plausible candidate for a moral virtue.

Given what we have said about the distinction between moral and epistemic virtues so far we take the following to be a plausible sufficient condition for moral virtue:

Moral Virtue: If a durable character trait non-accidentally disposes its possessors to act in ways that permissibly promote moral ends for the right reasons then it is a moral virtue.

We include ‘permissibly’ in the definition in order to prevent zealous devotion to one moral end that leads someone to perform forbidden acts from counting as a virtue. We say ‘for the right reasons’ to avoid the selfish person coming out as virtuous in a world where morality and self-interest converge.

Now we must ask ourselves whether moral wisdom meets our proposed sufficient condition. Remember that moral wisdom we took to be a disposition to possess true and only (or mainly) true moral beliefs. Given this moral wisdom does seem to be a trait that will dispose the agent to act in ways that permissibly promote moral ends for the right reasons. This is particularly true if we accept some form of Motivational Internalism about moral judgments. This is the view that moral beliefs are necessarily connected to motivation. If we accept some version of this view then it seems reasonable to think that moral wisdom is a trait that will lead people to promote moral

ends. Having correct moral beliefs will make the wise agent motivated to perform acts that promote moral ends. This in turn will increase the chances of the agent performing such acts, and non-accidentally so.

However, moral wisdom remains a plausible candidate for a moral virtue even if internalism is not true. Even motivational externalists accept that people's motivations are generally (rather than necessarily) in line with their moral beliefs. It is perhaps more controversial that these motivations are reliably "for the right reasons," but we think this optimistic hypothesis is very plausible insofar as we take common sense assessments at face value. For without this assumption, it will presumably be very rare that even virtuous moral agents act virtuously, since doing the right thing for the wrong reasons is paradigmatically failing to manifest virtue. This too can reasonably be contested, but we are here willing to trust common sense at least this far. This will be enough to make moral wisdom a moral virtue, as its possession will dispose agents to act in ways that promote moral ends.⁹

One worry we might have about our claim that moral wisdom is a moral virtue is that it is only plausible on a coarse grained account of moral wisdom. We have said that moral wisdom is the disposition to believe true and only true moral beliefs. However, it might be claimed that we should have a more fine-grained approach to the additional moral wisdom that Wise Inouye possesses. Rather than saying that Wise Inouye is more likely to possess true moral beliefs than Inouye we might say that Wise Inouye is more likely to accurately believe that an act is supererogatory rather than falsely believing it to be obligatory. When we describe Wise Inouye's additional moral wisdom in this way then it no longer seems as plausible to say that this is a moral virtue on our definition. After all, there seems no reason to think that a disposition to accurately class

⁹ See Doris (2002).

acts as supererogatory rather than obligatory will make Wise Inouye more disposed to promote moral ends. If anything we would expect this disposition to make Wise Inouye less likely to do so, since realizing that one's action is *not* morally required, which is entailed by its being supererogatory, may reduce one's motivation to perform it.

However, while it seems right that fine graining the disposition in this way will make it no longer appear to be a moral virtue, similar fine-graining worries might appear for any plausible candidate for a virtue. For instance, consider justice. Suppose one Police Officer possesses this virtue to a greater degree than another. We might think that this trait will make the just Police Officer more likely to arrest a friend who had committed a crime. When we say that she has a character trait that disposes her to treat everyone equally with respect to the law then it seems reasonable to say that this is promoting a moral end. However, we can fine grain this disposition in such a way that it no longer appears to be a virtue. A trait that makes one more disposed to arrest one's friends does not seem like a plausible candidate for a moral virtue. Similarly we might think a courageous person would be more likely to ignore risks to her own safety when in a position to save the lives of other people. This in turn will make her more likely to run into burning buildings than a less courageous person. Again, though, it is far from clear that being disposed to run into burning buildings promotes a moral end. While these points are far from decisive they give us some reason to be skeptical about the fine graining objection as it is applied exclusively to status of moral wisdom as a moral virtue.

Even overlooking the concern about fineness of grain, the objection can be met. For it turns out that knowing (or at least truly believing) that a given sort of action in a given sort of circumstance is not morally obligatory *does* promote a valuable moral end. For someone who has false positive beliefs about which actions are obligatory will be much

more prone to proselytize those not disposed to perform such actions, gossip about those who don't, and feel inappropriate resentment and/or indignation against those who do not. Such actions and attitudes are morally pernicious – they provide all of the trappings of a busybody. Immanuel Kant, often thought himself to be an overbearing moralist, is actually very good on this point:

But that man can be called **fantastically virtuous who allows nothing to be morally indifferent** and strews all his steps with duties, as with man-traps; it is not indifferent to him whether I eat meat or fish, drink beer or wine, supposing that both agree with me. Fantastic virtue is a concern with petty details which, **were it admitted into the doctrine of virtue, would turn the government of virtue into tyranny.** (Metaphysics of Morals, The Doctrine of Virtue, XVI)

We therefore conclude that moral wisdom where that wisdom is constituted by knowledge that a given action is *not* obligatory is indeed still a moral virtue. This does, though, leave us with another puzzle which we here only mention *en passant*. Actual moral heroes typically do *not* judge others harshly for failing to live up to their own standards precisely because they do not seem to universalize those standards. One possible explanation of this is that they are simply irrational in some way, and fail to understand the universality of fundamental moral principles. We think this is uncharitable, though, and that there is a more generous and interesting explanation of why people like Inouye do not view others who fail to act as they do in seemingly similar circumstances as thereby failing to do their duty.

Claim 3 ["No Other Relevant Difference"]: Wise Inouye possesses neither a greater level of some vice nor a lesser amount of some virtue than Inouye.

The third claim is that there are no vices that Wise Inouye possesses to a greater extent than Inouye. Similarly, there are no virtues that Inouye possess to a greater degree than Wise Inouye. This claim seems plausible when we consider the stipulated difference between Inouye and Wise Inouye. There seems no reason to think that by being a better judge of the moral status of his act Wise Inouye will thereby possess some virtue to a lesser degree or vice to a greater degree. Of course, the plausibility of this claim depends on the other differences that exist between Inouye and Wise Inouye. Again, though, there seems no obviously decisive reason to think that the closest possible world that meets our criteria will be one where Wise Inouye possesses a greater level of vice or lesser degree of virtue. The only *prima facie* plausible reason we can even think of for doubting Claim 3 would be that Wise Inouye is somehow lacking in modesty or humility. However, we suspect that this objection to Claim 3 rests on an implausible conception of those virtues or a misguided view of what Wise Inouye would have to be like. Still, this is an important worry about Claim 3 and we discuss it in more detail below.

In this section we have compared agents who instantiate the moral psychology described in section 1, which is apparently common amongst actual moral saints and heroes, with hypothetical agents who differ in correctly judging their actions to be supererogatory. We have made three intuitively plausible claims about the comparative judgements we would make about these two different agents. In the next section we argue that when we combine these three claims with a highly plausible principle about overall virtue assessment, a contradiction can be derived. Plausible theses lead through seemingly impeccable reasoning to a contradiction, so we will then have all of the ingredients of a paradox.

3. The Paradox

We are almost in a position to introduce our paradox. Before doing so we need to introduce a plausible principle about the way in which the virtues work.

Additional Virtue Principle (AVP): If there is no other difference in the virtues and vices possessed by two agents, x and y, then the fact that x possesses some virtue to a greater degree than y means that x is more virtuous than y.

There are two reasons to accept AVP. First, this just seems like good accounting. If we are adding together all of an agents virtues then it seems obvious that if we increase one of those virtues then the overall level of virtue we are left with will be greater.

Second, AVP seems like a plausible principle when we apply it to a relevant example. Imagine that the virtue that x possesses to a greater degree than y is the virtue of kindness. I think it is intuitively plausible to think that if two agents have exactly the same levels of virtue and vice apart from their levels of kindness then the kinder agent is more virtuous overall. These two reasons seem to be sufficient to establish the prima facie plausibility of AVP.

We are now in a position to set out the paradox:

- 1) Wise Inouye is not more morally virtuous overall than Inouye (and might even be less virtuous overall). (Claim 1).
- 2) Wise Inouye possesses more moral wisdom (itself a moral virtue) than Inouye. (Claim 2).
- 3) Wise Inouye possesses neither a greater level of some vice nor a lesser amount of some virtue than Inouye. (Claim 3).
- 4) If there is no other difference in the virtues and vices possessed by two agents, x and y, then the fact that x possesses some virtue to a greater degree than y means that x is more virtuous than y. (Additional Virtue Principle).

As we have already established, these four statements are all plausible when considered independently of one another. When we consider them together, though, they are inconsistent. If AVP is true then Claim 1 contradicts Claims 2 and 3.

4. Resolving the Paradox

In our view, Claim 2 [More Moral Wisdom] and AVP are very hard to deny. AVP just seems like a platitude about overall virtue, and Claim 2 seems to follow from the description of the case plus the thesis that, *ceteris paribus*, more substantive moral knowledge means more moral wisdom. How much more moral wisdom may depend on how significant the additional moral propositions known are, but knowing whether it is one's moral duty to perform some action which involves a huge sacrifice strikes us as a pretty significant sort of moral knowledge, especially if the circumstances in which the choice to perform the action are not entirely hypothetical.

If this is right then to avoid the looming contradiction we should either give up Claim 1 [Not More Virtuous] or Claim 3 [No Other Relevant Difference] (the 'or' is inclusive). Giving up Not More Virtuous might seem like the more promising option. After all, Not More Virtuous relies on a direct appeal to our intuitions about a counterfactual, but it may simply be unclear just what the nearest possible world in which Wise Inouye has a more accurate view of his moral duties is like. Whether Wise Inouye is more virtuous than Inouye is going to depend on what *other* differences we find between the actual world and the nearest possible world(s) in question. We initially assumed that there was a single unique closest possible world in which Inouye does not make his mistake, but this is of course unrealistic. More realistically, there may well be a sort of tie, with a wide range of equally modally "distant" worlds in which Wise Inouye is as we describe him. This may further muddy the waters, if in some of those worlds Wise

Inouye is more virtuous than Inouye while in others he is not. Given that it is in this way just not clear what we can and cannot legitimately assume about the relevant counterfactual, perhaps our intuitions in support of Not More Virtuous are simply unreliable, and we should happily reject that thesis to avoid the paradox.

We have to be careful here, though. Claims 2, 3 and AVP entail the negation of Not More Virtuous. So if it turned out that Not More Virtuous were indeterminate because in some of the nearest worlds Wise Inouye is more virtuous but in some he is not (and there is no clear preponderance, say) then this would not resolve our paradox. For the other theses which make up the paradox entail that Not More Virtuous is simply *false*, not indeterminate in truth value.

In our view, a more promising strategy is to put pressure on No Other Relevant Difference. We begin, though, with explaining why one strategy for rejecting No Other Relevant Difference strikes us as unhelpful. One might argue that Wise Inouye is somehow smug, or less modest or humble, than Inouye, and that this is a relevant difference which explains why he is not overall more virtuous. One might imagine Wise Inouye boasting about how he went above and beyond the call of duty, patting himself on the back, and the like.

This, though, rests on a gratuitously uncharitable view of what the nearest world in which a moral hero like Inouye has a more accurate view of his moral duties would be like. More plausibly, while Wise Inouye would know he went above and beyond the call of duty, he would feel no need to “broadcast” this to the world. Indeed, he might be modest enough to try to avoid the topic, and only confess to having gone above and beyond if pressed in an interview – and in that scenario the demands of honesty might force him to allow that he did go above and beyond the call of duty. This, though,

hardly seems to make him come out as lacking in modesty, at least not in any sense in which modesty is a virtue.¹⁰

Perhaps, though, there is a different way of supporting the claim that Inouye is more humble than Wise Inouye. We might think that what explains Inouye's claim that he performed an obligation is an admirable disposition to err on the side of moral safety.¹¹

It seems reasonable to think that when unsure about whether an available act is obligatory or supererogatory, the virtuous person will take the morally safe option and perform the act. Perhaps, Inouye is displaying a similar virtue, when he is unsure about whether an act available to him is obligatory or supererogatory he works on the assumption that it is obligatory in order to err on the side of moral safety. Given this, we might think that the morally relevant difference between Inouye and Wise Inouye is that Inouye has been humble enough to recognise the possibility of making a mistaken moral judgement in this situation and, as a result, has worked on the assumption that the act is obligatory. Wise Inouye, on the other hand, has not been so humble, as he has been quite happy to accept his judgement that the act was obligatory.

However, this response is problematic given the existing empirical evidence about moral heroes like Inouye. Someone who is unsure about the moral status of his act would be unlikely to explain his action by claiming without qualification that the reason he acted as he did was because he had a duty to do so, at least not if he is honest. This explanation also clashes with the observations made by those studying the psychology of moral exemplars. One of the observations made by Colby and Damon in their psychological study of moral exemplars, many of whom provide examples of the phenomenon we are discussing, is the *certainty* with which they act. In their words,

¹⁰ For one of the authors' views of how we should understand modesty as a virtue, see [Author's Citation Removed]

¹¹ Thanks to an anonymous referee for suggesting this response.

They are people who translate their principles into action directly, with little indecision or hesitation. There is a great sense of certainty, and a conspicuous absence of doubt, in their moral conduct (1992: 70).

If Colby and Damon are right about this then it does not seem plausible to explain Inouye's action by appealing to a desire to err on the side of moral safety. More generally, moral heroes seem to be quite confident that they are obligated to act as they do. So although there are possible cases which have this structure, of course (and perhaps some actual cases too), this move ultimately does not seem to provide a satisfactory explanation of the core puzzle cases.

Still, we think No Other Relevant Difference is the weakest member of the set of theses that yields our paradox, but for a rather different reason. To see why No Other Relevant Difference can reasonably be rejected, it will be useful to survey some of the empirical evidence about the moral psychology of actual moral heroes like Inouye.

Anne Colby and William Damon have done interesting work in this regard. They conducted in-depth case studies of 23 moral exemplars of the relevant sort and tried to see what these figures had in common in virtue of which they acted so heroically. The most relevant finding, for our purposes, is that these moral heroes did not seem to consider their actions heroic precisely because they *identified* so closely with the relevant moral values that they saw no gap between doing whatever was necessary to promote or respect those values, and their own self-interest. Here are Colby and Damon:

The exemplars have done so without devaluing their own personal goals. Nor do they disregard their own fulfillment or self-development, nor, broadly

construed, their own self-interests. They do not seek martyrdom. Rather than denying the self, they define it with a moral centre. (Colby and Damon: 300)

The point Colby and Damon are making is that we should not think of exemplars as acting against their own goals. Rather, exemplars make moral goals a key part of their identity. According to Colby and Damon, the identification of self-interest with moral goals is not something that exemplars have innately. Rather it is a gradual process by which the moral goals people strive towards become slowly harder to distinguish from the exemplars self-interested goals.¹²

While Colby and Damon's study is based on a small sample size their findings are supported by two later studies of moral exemplars. Frimer et. al. (2011). They found that, compared to the comparison group, exemplars were significantly more likely to have integrated their own interests with their moral interests (2011 p.160). A study of influential historical figures by Frimer et al. (2012) found similar results.

Suppose this is right, that moral heroes do tend to identify with their values in this way. One question that immediately arises is whether in this context "thinking makes it so." That is, perhaps on some views of human welfare, if one identifies wholeheartedly enough with a value then it really is true that one's welfare and responding appropriately to that value cannot come apart. If this were true, then, somewhat surprisingly, our moral heroes would not have been making a sacrifice in the seemingly relevant sense of making themselves worse off in acting as they did. In fact, they would have been making a sacrifice if they had acted otherwise!

¹² See Colby and Damon (1992 Chapter 7) and Frimer and Walker (2009) for explanations of the development of moral commitment in exemplars.

We find this conclusion implausible, and so we reject the thesis that “thinking makes it so” in this strong sense in this case. Note though, that if we are wrong about this then another resolution of our paradox seems to open up. For now it may turn out that actual moral heroes (assuming they are still aptly so-called, on this view of their welfare!) are not mistaken about having only done their duty. After all, they did great good and at no sacrifice to themselves, all things considered – isn’t that enough to make their actions obligatory after all?¹³ If, though, they are right about that, then More Moral Wisdom [claim 2] is false – Inouye and his ilk are not mistaken, after all, about their duties.

We find this line implausible and not only because we reject this very strong theory of the connection between the values with which one identifies and one’s overall self-interest. The paradox remains because there will still be at least some possible cases in which the moral hero does not identify wholeheartedly enough with the relevant value to make it come out, even on such a high-minded theory of welfare, that they did not make a sacrifice in acting as they did.

Still, it seems highly relevant to the resolution of our paradox that moral heroes generally, identify with morality in this strong way, and that this explains *why* they are inclined to say that they simply did what they had to do – that they were just doing their duty. For this level of identification with morality, or with some specific moral value, strikes us as an important moral virtue – a form of moral *depth*.

¹³ In fact, it isn’t clear that this would make these acts obligatory. A number of philosophers have argued that it is possible for an act of supererogation to be inline with the agents self interest. See Kawall (2003), Ferry (2013; 579) and Horgan and Timmons (2010: 54). For one of the author’s argument against the claim that supererogation always involves sacrifice see [Author’s Citation Removed].

This suggests a plausible explanation for the differing views that Inouye and Wise Inouye have of their actions. Perhaps in the nearest possible world in which Wise Inouye has a more accurate view of his moral duties this is because his identification with the relevant moral value(s) (in his case, perhaps solidarity with his fellow soldiers, a sense of patriotic duty informed by gratitude for what his country had done for him) is less wholehearted. While Inouye would find it hard to distinguish his own goals from that of his regiment, these are clearly distinct for Wise Inouye.

That is not to say that Wise Inouye does not value the goals of the regiment. Rather the point is that he can clearly see the difference between these two conflicting values and how they can come apart, while this is, at the very least, much less clear for Inouye. This, then, means that Wise Inouye buys his greater moral wisdom at a cost – the cost of less *moral depth*. In that case, we have plausible grounds for rejecting No Other Relevant Difference, and one which does not rely on an uncharitable view of Wise Inouye as smug or immodest. Note that on this solution to the paradox the compensating virtue is strongly connected to the lesser moral wisdom. It is Wise Inouye's identification with his moral commitments that makes him pre-disposed to have a skewed assessment of his own interests.

Of course, for this solution to the paradox to get off the ground we must say something about what moral depth is and why Inouye's identification with his moral goals suggests that he possesses a greater degree of moral moral depth than Wise Inouye. First, then, we must say what we mean by moral depth. As it happens, not a lot of work has been done in analytic philosophy about the nature of moral depth, and we will here only be able to scratch the surface of this independently very important and neglected topic.

One of the few papers in analytic philosophy that discusses moral depth is by John Kekes (in his “Moral Depth”). Kekes begins by giving the following account of depth in general:

Depth involves discerning an underlying unity among apparently complex and unrelated phenomena. It is to see the same phenomena as many others also see, but to penetrate below the surface and construct a theory or a vision, depending on the subject-matter, that leads to a possible understanding of the reality of which the appearances are manifestations (1990: 440).

Kekes argues that depth has three related aspects – depth of understanding, emotional depth and a sort of practical depth which is more concerned directly with the guidance of action. Kekes further argues that depth of understanding is in a sense primary, in that so-called emotional or practical depth not based on a proper understanding of the relevant norms would not really be worthy of being classified as forms of true depth. As Kekes memorably puts the point, depth of understanding is essential to distinguishing true depth from mere “humbug.” (Kekes 1990: 442) Depth of understanding is, moreover, not simply knowing some related propositions. It is instead a matter of seeing an underlying unity among seemingly unrelated phenomena.

Having explained what depth is in general, Kekes then goes on to give an account of *moral* depth. According to Kekes, the subject matter of moral depth is the question of how to live a good life. To possess moral depth then, is to possess a vision of what it is to lead a good life. This might seem to pose a challenge to our view, since moral heroes like Inouye are in some ways strikingly lacking in an adequate moral understanding. In particular, of course, they manifest a failure to distinguish moral duty from supererogation in the right way.

We think this objection is misguided, though, and for two reasons. First, although moral heroes do characteristically go wrong in this way, their emotional and practical moral depth is not *based on* this mistake. Rather, it is the other way around, or so we have argued. It is their deep emotional and practical identification with the relevant moral values which primes them for failing to mark the duty/supererogation boundary in precisely the right way. Moreover, the depth of emotional resonance with the relevant moral values and the practical depth of their commitment to uphold those values itself reflects a kind of deep moral understanding. A paradigmatic moral hero seems to recognize (more than most of us) the important moral truth that he is but one person in a much wider moral universe, and that he ultimately counts for no more than anyone else.

One could, of course, argue independently of Kekes' thesis for the stronger view that simply because these aspects of their character primes them for this sort of mistake that the aspects in question should not qualify as constituting moral depth. This, though, seems far too heavy handed to us. Inouye and other moral heroes are not plausibly accused of being purveyors of humbug. More generally, if the mere fact that some character trait made it more likely that one might get some specific moral question wrong prevented that trait from being a virtue then a great many otherwise plausible instances of moral virtue may be disqualified far too quickly. Loyalty, for example, may prime one to overlook the faults of one's friends (for closely related discussion, see Driver 2001).

Second, we are not as convinced as Kekes that depth of moral understanding is essential for emotional and practical aspects of moral depth. Granted, emotional and practical aspects of moral depth must not be based on *misunderstandings* of the relevant moral values, but that is not the case for Inouye and others. The depth of their

commitments springs from an appreciation of the significance of the relevant moral value and a deep recognition of the fundamental equality of all moral agents. Kekes' arguments make it very plausible that these other forms of depth must not be based on a kind of moral confusion, but we think it is a big leap from that to the much stronger conclusion that one can manifest emotional and practical forms of moral depth *only if* one manifests exceptional moral understanding. Pre-theoretically, it strikes us as plausible that many people who we might describe as "salt of the earth" characters with strong moral convictions can manifest practical/emotional moral depth without being especially above average in their appreciation of the ways in which seemingly diverse moral phenomena can be unified. Colby and Damon make a relevant point in the following:

Pondering moral problems is not the same as dedicating one's life to their solution. The capacity for single-minded dedication to a moral cause may have little to do with the capacity to reason about abstract moral principles. The will to take a stand may derive from a source entirely different from the ability to arrive at sophisticated intellectual judgement (1994: 6).

Plausibly, someone who dedicates their life to the solution to moral problems is just as deserving of being described as possessing moral depth as someone who devotes her life to the solution of these problems. What these agents may lack in their capacity for abstract reasoning about moral principles they more than make up in the depth of their practical and emotional appreciation of what it is to lead a good life. The insistence that such depth of moral understanding is essential for *anything* worthy of the label 'moral depth', then, strikes us as an overly intellectual conception of moral depth more generally. Granted, depth of moral understanding is *one* important aspect of moral depth, but it is plausibly not the only such aspect.

Indeed, formulating the point in terms of different “aspects” of moral depth implicitly agrees with a further thesis put forward by Kekes, namely that moral depth is a single phenomenon with many different aspects, rather than being a genus with a number of different but related species. Although it is not clear that our proposal here relies essentially upon it, we actually think it is much more plausible to think that moral depth does indeed come in different forms, but that these forms are still recognizably species of the same genus. Of course, it is a good question just what unifies these different species under a single non-gerrymandered genus. A full account of this matter would unfortunately take us too far afield, though, and would require a fully worked out theory of moral depth. We suspect, though, that these different forms of depth all function to reinforce one another in interesting ways, thus providing a virtuous sort of homeostatic cluster. We cannot do anything to substantiate this suspicion within the confines of this paper.

However, with these important disagreements out of the way, we find much to agree with in Kekes’ discussion, and indeed much which corroborates our hypothesis about Inouye and company. In particular, Kekes’ plausible conception of the emotional and practical aspects (or, as we might prefer, *forms*) of moral depth actually fits very well with our hypothesis about moral heroes like Inouye. For on Kekes’ account, these aspects of moral depth constitutively involve the harmonization of moral virtue and self-interest. Here is a key passage from Kekes:

Moral depth is important, because it involves understanding the significance of the relevant conditions for the human aspiration to lead good lives; lives, that is, in which there is a great deal of personal satisfaction and moral merit, and in which the two are so related that the achievement of one coincides with the achievement of the other. (Kekes 1990: 445)

This strikes us as a very plausible conception of moral depth, at least in its emotional and practical aspects (or forms).

This might seem to conflict with one of our earlier claims, namely that moral heroes are mistaken in thinking that their actions are obligatory because they involve no sacrifice. How can this be reconciled with our endorsement of Kekes' view of moral depth, as at least partially being constituted by finding ways to make morality and personal satisfaction coincide? Insofar as this really reflects moral depth, then it seems like it must be true that morality and self-interest converge for the ideally virtuous person. In that case, though, the tight connection between promoting the relevant moral value(s) and leading a good life presupposed by moral heroes is sound, after all, in which case it is less obvious why their actions are supererogatory as opposed to obligatory.

The answer to this challenge is that there are limits on the extent to which the pursuit of moral ends can constitute one's welfare. If the sacrifice in terms of pain and suffering, lifelong physical and mental debilitation, loss of one's rational capacities or autonomy is great enough then the fact that one has promoted a cherished moral value does not mean one has not on balance made a sacrifice, indeed a large sacrifice. This is, we think, compatible with granting that it is a genuine and deeply important insight that morality can *partly* constitute one's welfare when one identifies with it, and with Kekes' idea that moral depth in part is a matter of recognizing this and managing to identify with morality in the right way so that one does have this convergence. Nonetheless, we think that, plausibly, morality will still only partly constitute one's welfare and that this component of a person's welfare will still not take lexical priority over other non-moral goods. So while moral heroes do exhibit very real and exceptional moral depth in their identification with the relevant moral values, they also typically get so carried away by their enthusiasm for those values that they fail to recognize their own very real

sacrifices, and thereby mistake what is actually supererogatory for a moral obligation.

We do not think this further quite understandable lapse in moral wisdom should disqualify them from exhibiting the relevant sort of moral depth in virtue of their practical and affective commitments, though, for reasons we have outlined above.

Moreover, this conception of moral depth makes it clear why moral depth should be considered a moral virtue. In Section One we claimed that if a durable character trait non-accidentally disposes its possessors to act in ways that permissibly promote moral ends for the right reasons then it is a moral virtue. An agent who receives personal satisfaction from the achievement of moral goals will be less inclined to view the best act from the moral point of view as in conflict with the best act from a self-interested point of view. This lack of conflict will make the agent less disposed to letting considerations of self-interest get in the way of performing acts that permissibly promote moral ends for the right reasons.

Of course, it is hard to imagine clearer instances of such harmonization than one finds in the lives of moral heroes like Inouye, and as described by Colby and Damon. So if anything like Kekes' account is correct, the hypothesis that Inouye and company manifest great moral depth in virtue of their identification of self-interest and the relevant moral values is very plausible indeed. This, in turn, quite elegantly explains why Inouye is not obviously less morally virtuous than Wise Inouye. For plausibly, in the nearest possible worlds in which Inouye realizes that he is not merely doing his duty he does so in part because his identification with the relevant moral values is less wholehearted and essential to his conception of his own identity and welfare.

Conclusion

We began this paper by investigating a common phenomenon amongst those who perform supererogatory acts. Often those who perform such acts claim it was their duty to do so. We argued that a paradox can be created by combining three independently plausible but jointly inconsistent claims about how we should view someone who makes this mistaken judgement compared to someone who does not. We argued that the best way to resolve this paradox is to appeal to the virtue of moral depth that the former may plausibly be said to possess to a greater degree than the latter. This virtue predisposes moral exemplars to make this mistake about the deontic status of their acts and provides the best explanation of the ubiquity of this mistake amongst moral heroes. This explanation fits well with the empirical data, and also provides an independently plausible explanation of why moral heroes seem not to hold others to the same lofty standards that they impose on themselves. Happily, the explanation also resolves our paradox by explaining why Inouye is not less virtuous than Wise Inouye, namely because he has greater moral depth.

Our analysis of the phenomenon also provides a partial vindication for Julia Driver's claims, mentioned in passing in Section One, about the relationship between moral virtue and ignorance. We have claimed that there is a moral virtue that disposes the possessor to a characteristic sort of error. This does not go as far as Driver's claim that some moral virtues (such as modesty, blind charity and trust) are partly constituted by ignorance (2001: Ch. 2) though, while we are not committed to Driver's stronger thesis, nothing we say is incompatible with it. However, our weaker claim does provide some support for Driver's wider point that an acceptable theory of the moral virtues ought to allow that the morally virtuous agent might make epistemic mistakes. After all, it would be odd if a virtue that is typically possessed by moral exemplars and that disposes the possessor to perform the right act for the right reasons

were also to prevent the possessor from counting as virtuous because it disposes her to make epistemic mistakes.

Bibliography

Boghani, Priyanka (2012) "Sen. Daniel Inouye, War Hero" In *Global Post* (18/12/12)

Available at: <http://www.globalpost.com/dispatch/news/regions/americas/united-states/121218/sen-daniel-inouye-war-hero-video> (Accessed 20/12/12).

Buckley, Cara (2007) "Man is Rescued by Stranger on Subway Tracks" In *New York Times* (3/1/2007) Available at:

<http://www.nytimes.com/2007/01/03/nyregion/03life.html?ex=1325480400&en=bf239e4fab06ab5&ei=5090&partner=rssuserland&> (Accessed 20/12/12).

Carbonell, V. (2012) "The Ratcheting-up Effect" In *Pacific Philosophical Quarterly* Vol. 93 pp. 228-254.

Colby, A. and Damon, W. 1992. *Some Do Care*. Free Press.

Doris, J. (2002) *Lack of Character: Personality and Moral Behaviour*. (New York: Cambridge University Press)

Driver, J. (2001) *Uneasy Virtue* (Cambridge: Cambridge University Press).

_____ (2003) "The Conflation of Moral and Epistemic Virtue", in *Metaphilosophy* (Vol. 34 No. 3 pp.367 - 383).

Ferry, M. (2013) "Does Morality Demand Our Very Best? Moral Prescriptions and the Line of Duty" In *Philosophical Studies* 165 (2):573-589.

Frimer, J. A., & Walker, L. J. (2009). "Reconciling the self and morality: An empirical model of moral centrality development." *Developmental Psychology*, 45, 1669-1681.

- Frimer, J. A., Walker, L. J., Dunlop, W. L., Lee, B. H., & Riches, A. (2011). "The integration of agency and communion in moral personality: Evidence of enlightened self-interest." *Journal of Personality and Social Psychology*, 101, 149-163.
- Frimer, J. A., Walker, L. J., Lee, B. H., Riches, A., & Dunlop, W. L. (2012). Hierarchical integration of agency and communion: A study of influential moral figures. *Journal of Personality*, 80(4), 1117-1145.
- Gomez-Granger, Julissa (2008) *CRS Report for Congress: Medal of Honor Recipients: 1979-2008* (RL30011 4/6/2008) Open Crs (Available From: <https://opencrs.com/document/RL30011>/Accessed 20/12/12)
- Heyd, David (2011) "Supererogation", in *The Stanford Encyclopedia of Philosophy* (Winter 2011 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2011/entries/supererogation/>
- Horgan, T. and Timmons, M., (2010), "Untying a Knot from the Inside Out: Reflections on the 'Paradox' of Supererogation", In *Social Philosophy and Policy*, 27: 29-63.
- Kawall, J. (2003). "Self-Regarding Supererogatory Acts" in *Journal of Social Philosophy* 34 (3): 487-498.
- Kekes, J. (1990) "Moral Depth," *Philosophy* (pp. 439-453).
- Montmarquet, J. (1986). "Epistemic Virtue." In *Mind*, (pp. 482-97).
- Mulderink, Earl F. III (2012) *New Bedford's Civil War* (New York: Fordham University Press)
- Portmore, D. (2008). "Are moral reasons morally overriding?" In *Ethical Theory and Moral Practice* 11: 369-388.

Zagzebski, L. (1996). *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*. (Cambridge: Cambridge University Press).